

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

Abridged Translation of Japanese Laid-Open Patent

Publication No. 9-233442

(Published on September 5, 1997)

Japanese Patent Application No. 8-033956

(Filed on February 21, 1996)

Title: FACE IMAGE DISPLAY DEVICE AND FACE IMAGE
COMMUNICATION SYSTEM

Applicant: CASIO COMPUTER CO., LTD.

(Paragraphs 43 and 48)

[0043]

For example, 20 types of message frame pattern data P01 through P20 (part number) are stored in the message frame pattern ROM 31 as shown in FIG. 3. Each of the message frame pattern data P01 through P20 corresponds to a number of identified characters as described later on. The message frame pattern, i.e., an image of message screen is created based on the message frame pattern data. The message frame pattern comprises a matrix of $n \times m$ dots corresponding to the number of characters in a phrase of a message, i.e., the number of characters in a phrase identified by a sound identification unit 10 as described later on. For example, if three characters are identified, a message frame pattern of $N3 \times M3$ dots is selected based on the message frame pattern data P03.

[0048]

The message frame image RAM 42 is used to combine image patterns (the message frame pattern and the message pattern) to make a composite pattern in a

received image, at an area other than a specific area dedicated to an image of a person, i.e., to place the message pattern in the message frame pattern. The composition RAM 43 is used to convert a received image into a composite image of dot data, i.e., to create an image of the composite pattern for the area other than the specific area created by the message frame image RAM 42. The composite image of dot data created by the composition RAM 43 is used as display data for displaying an image by the display unit 13.



PATENT ABSTRACTS OF JAPAN

(11) Publication number: **09233442 A**(43) Date of publication of application: **05.09.97**

(51) Int. Cl. **H04N 7/14**
G10L 3/00
G10L 3/00
G10L 3/00
H04M 11/06

(21) Application number: **08033956**(22) Date of filing: **21.02.96**(71) Applicant: **CASIO COMPUT CO LTD**(72) Inventor: **MURATA YOSHIYUKI**

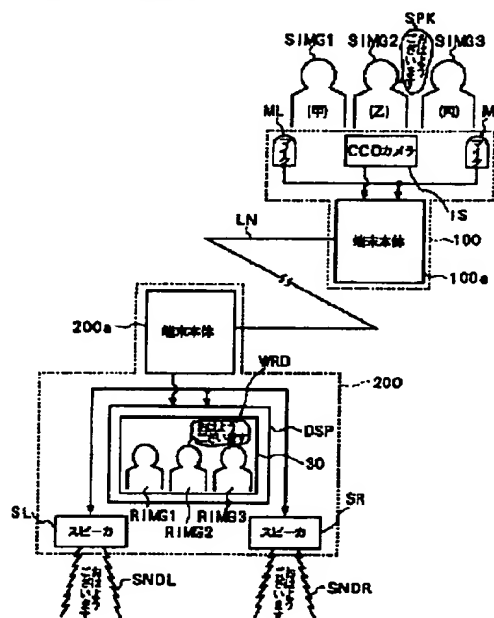
(54) **FACE IMAGE DISPLAY DEVICE AND FACE
 IMAGE COMMUNICATION SYSTEM**

COPYRIGHT: (C)1997,JPO

(57) Abstract:

PROBLEM TO BE SOLVED: To surely and easily visually confirm voice contents uttered by a person or an animal just under display even by a person handicapped in hearing.

SOLUTION: While using an image input part having a CCD camera IS and a voice input part having microphones ML and MR, information communication equipment 100 on the transmission side inputs an image and a voice and sends them to a line LN as communication data. Based on image data in the communication data received from the line LN, information communication equipment 200 on the reception side detects any prescribed change area (mouth) and based on voice data in these communication data, the voice is recognized. Then, the group of characters corresponding to the voice recognized near the prescribed change area (mouth) is displayed in the received image on a display screen 30, and that voice is discharged from speakers SL and SR in stereo state.



(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-233442

(43)公開日 平成9年(1997)9月5日

(51)Int.Cl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
H 0 4 N 7/14			H 0 4 N 7/14	
G 1 0 L 3/00	5 1 3		G 1 0 L 3/00	5 1 3 Z
	5 6 1			5 6 1 C
	5 7 1			5 7 1 G
H 0 4 M 11/06			H 0 4 M 11/06	
審査請求 未請求 請求項の数9 O L (全 18 頁)				

(21)出願番号 特願平8-33956

(22)出願日 平成8年(1996)2月21日

(71)出願人 000001443

カシオ計算機株式会社

東京都新宿区西新宿2丁目6番1号

(72)発明者 村田 嘉行

東京都羽村市栄町3丁目2番1号 カシオ
計算機株式会社羽村技術センター内

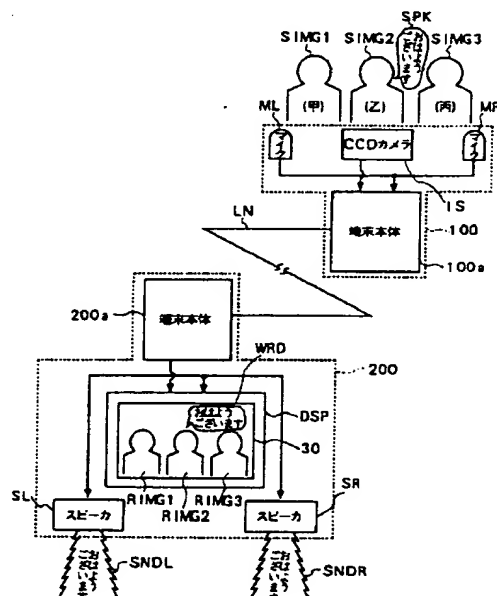
(74)代理人 弁理士 荒船 博司 (外1名)

(54)【発明の名称】 顔画像表示装置及び顔画像通信システム

(57)【要約】

【課題】 耳の不自由な人にとっても、表示中の人物や動物等が発している音声内容を確実かつ容易に視認できるようにすることを課題とする。

【解決手段】 送信側の情報通信装置100は、CCDカメラISをもつ画像入力部及びマイクML, MRをもつ音声入力部により画像及び音声を入力し、これを通信データとして回線LNに送出し、受信側の情報通信装置200は、回線LNより受信された通信データの内の画像データに基づいて所定の変化領域(口元)を検出するとともに、その通信データの内の音声データに基づいて音声を認識して、表示画面30の受信画像中に所定の変化領域(口元)近傍に認識された音声対応の文字群を表示するとともに、その音声をスピーカSL, SRよりステレオ状態で放音する。



【特許請求の範囲】

【請求項1】外部から送られてきた顔画像データ及び音声データを含むデータを受信する受信手段と、前記受信手段により受信された顔画像データに基づいて対応する顔画像を表示する表示手段と、前記受信手段により受信されたデータの内の顔画像データに基づいて口元の変化領域を検出する検出手段と、前記受信手段により受信されたデータの内の音声データに基づいて音声を認識する音声認識手段と、前記音声認識手段により認識された音声に対応する文字を出力する文字出力手段と、前記表示手段により表示されている顔画像の内、前記検出手段により検出された前記口元の変化領域の位置又はその近傍位置に前記文字出力手段から出力された音声に対応する文字を表示制御する表示制御手段と、を有することを特徴とする顔画像表示装置。

【請求項2】外部から送られてきた顔画像データ及び複数ヶ所より集音された複数の音声データを含むデータを受信する受信手段と、前記受信手段により受信されたデータの内の顔画像データに基づいて対応する顔画像を表示する表示手段と、前記受信手段により受信されたデータの内の顔画像データに基づいて口元の変化領域を検出する口元位置検出手段と、前記受信手段により受信されたデータの内の音声データに基づいて音声を認識する音声認識手段と、前記音声認識手段により認識された音声に対応する文字を出力する文字出力手段と、前記受信手段により受信されたデータの内の複数の音声データに基づいて当該複数の音声データそれぞれの発生位置を検出する音声位置検出手段と、前記表示手段に表示されている顔画像の内、前記口元位置検出手段により検出された口元の変化領域の位置又はその近傍位置と前記音声位置検出手段により検出された各音声の発生位置に対応する表示位置との少なくとも一方の位置に、前記文字出力手段から出力された音声に対応する文字を表示制御する表示制御手段と、を有することを特徴とする顔画像表示装置。

【請求項3】前記受信手段により受信されたデータの内の音声データに基づいて対応する音声を外部に出力する音声出力手段を更に有することを特徴とする請求項1又は2に記載の顔画像表示装置。

【請求項4】前記文字出力手段から出力される音声に対応する文字を包囲する吹出し画像が記憶されている吹出し記憶手段とを更に備えており、前記表示制御手段は前記文字出力手段から文字を出力した際に、その文字の表示領域を包囲する吹出し画像を前記吹出し記憶手段から読み出して合成表示する吹出し表示制御手段を含むことを特徴とする請求項1又は2に記載の顔画像表示装置。

【請求項5】前記吹出し記憶手段は、

表示文字数に対応させてサイズの異なる複数の吹出し画像を記憶する吹出し画像記憶手段を備えており、

前記表示制御手段は、

前記文字出力手段から出力された音声に対応する文字の文字数に対応する吹出し画像を前記吹出し記憶手段に記憶された複数の吹出し画像の中から読み出して表示する読出表示手段を備えていることを特徴とする請求項4記載の顔画像表示装置。

【請求項6】さらに、

10 前記受信手段により受信されたデータの内の顔画像データに基づいて前記表示手段の表示領域の内の特定領域を検出する特定領域検出手段を有し、

前記吹出し表示制御手段は、前記受信手段により受信されたデータの内の顔画像データに基づく表示画面上で前記特定領域検出手段により検出された特定領域の表示領域とは別の領域に前記吹出し画像を表示することを特徴とする請求項4又は5に記載の顔画像表示装置。

【請求項7】前記文字出力手段から出力される音声に対応する文字を包囲する吹出し画像およびこの吹出し画像と前記口元位置検出手段により検出された口元位置とに配置される吹出し起点画像がそれぞれ記憶されている画像記憶手段とを更に備えており、前記表示制御手段は前記文字出力手段から文字を出力した際に、その文字の表示領域を包囲する吹出し画像と前記口元位置検出手段により検出された口元位置に配置される吹出し起点画像とを前記画像記憶手段から読み出して合成表示する吹出し表示手段を含むことを特徴とする請求項3記載の顔画像表示装置。

【請求項8】請求項1記載の顔画像表示装置と、前記顔画像表示装置に対し回線を介してデータを送信する送信装置とからなる顔画像通信システムにおいて、

前記送信装置は、

顔画像データを入力する画像入力手段と、

音声データを入力する音声入力手段と、

前記画像入力手段により入力された顔画像データと前記音声入力手段により入力された音声データとをデータとして前記顔画像表示装置に送信する送信手段と、

を有することを特徴とする顔画像通信システム。

【請求項9】請求項2記載の顔画像表示装置と、前記顔画像表示装置に対し回線を介してデータを送信する送信装置とからなる顔画像通信システムにおいて、

前記送信装置は、

複数の顔画像データを入力する画像入力手段と、

複数ヶ所より集音された複数の音声データを入力する音声入力手段と、

前記画像入力手段により入力された複数の顔画像データと前記音声入力手段により入力された複数の音声データとをデータとして前記顔画像表示装置に送信する送信手段と、

50 有することを特徴とする顔画像通信システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、外部から送られてきた顔画像を表示する顔画像表示装置、及びその顔画像表示装置に対して送信装置から回線を介してデータを送信する顔画像通信システムに関する。

【0002】

【従来の技術】従来、この種の顔画像通信システムには、代表例として、テレビ電話方式を採用したものがある。

【0003】このテレビ電話方式は、顔画像表示装置である送信側端末にカメラをセットし、そのカメラによって撮影した人物像（画像のひとつ）及び会話（音声のひとつ）を画像データ及び音声データとして通信回線を介して顔画像表示装置である受信側端末に送信される仕組みである。この受信側端末では、受信された画像データ及び音声データに基づいてCRT等の画像出力装置に表示及び放音出力が行われ、遠隔地に送信側端末の音声付き映像を取得することができる。

【0004】また、このテレビ電話方式では、上記受信側端末から上記送信側端末に対して画像データ及び音声データを送信して上記送信側端末で上記受信側端末の音声付き映像を出得することもできることから、双方向通信を可能にしている。

【0005】

【発明が解決しようとする課題】ところで、上述した従来例による顔画像通信システムにおいては、人物像の表示出力とともにそのときの会話を放音出力するようにしたので、遠隔地にいる相手の表情や動きを表示画像から読み取って電話だけのコミュニケーションよりも相手に伝達できる情報の種類を増した分だけ正確にコミュニケーションすることが可能であるが、耳の不自由な人にとってはもとより、この伝達する情報の種類を増やずに、テレビの文字放送のように放音だけでなく文字の表示出力を伴って会話することのできる簡易機能に期待がもたれている。

【0006】本発明の目的は、上述した従来例による課題を解消するため、耳の不自由な人にとっても表示中の人物や動物等が発している音声内容を確実かつ容易に視認することが可能な顔画像表示装置を提供することにある。

【0007】本発明の他の目的は、簡単なシステム構成で音声に対応する文字と顔画像とを表示出力することが可能な顔画像通信システムを提供することにある。

【0008】

【課題を解決するための手段】請求項1記載の発明に係る顔画像表示装置は、外部から送られてきた顔画像データ及び音声データを含むデータを受信する受信手段と、前記受信手段により受信された顔画像データに基づいて対応する顔画像を表示する表示手段と、前記受信手段に

より受信されたデータの内の顔画像データに基づいて口元の変化領域を検出する検出手段と、前記受信手段により受信されたデータの内の音声データに基づいて音声を認識する音声認識手段と、前記音声認識手段により認識された音声に対応する文字を出力する文字出力手段と、前記表示手段により表示されている顔画像の内、前記検出手段により検出された前記口元の変化領域の位置又はその近傍位置に前記文字出力手段から出力された音声に対応する文字を表示制御する表示制御手段と、を有することを特徴とする。

【0009】以上の構成によれば、受信手段は外部から送られてきた顔画像データ及び音声データを含むデータを受信し、表示手段は受信手段により受信された顔画像データに基づいて対応する顔画像を表示し、検出手段は受信手段により受信されたデータの内の顔画像データに基づいて口元の変化領域を検出し、音声認識手段は受信手段により受信されたデータの内の音声データに基づいて音声を認識し、文字出力手段は音声認識手段により認識された音声に対応する文字を出力し、表示制御手段は、表示手段により表示されている顔画像の内、検出手段により検出された口元の変化領域の位置又はその近傍位置に文字出力手段から出力された音声に対応する文字を表示制御する。

【0010】従って、受信された顔画像データに基づく顔画像を表示出力する際にその顔画像の口元の位置又はその近傍位置に、受信された音声対応の文字を同時に表示するようにしたので、複数の顔画像から同時に音声対応の文字が表示された場合でも、その各文字と複数の顔画像との対応関係を明確に視認することが可能となる。

【0011】請求項2記載の発明に係る顔画像表示装置は、外部から送られてきた顔画像データ及び複数の所より集音された複数の音声データを含むデータを受信する受信手段と、前記受信手段により受信されたデータの内の顔画像データに基づいて対応する顔画像を表示する表示手段と、前記受信手段により受信されたデータの内の顔画像データに基づいて口元の変化領域を検出する口元位置検出手段と、前記受信手段により受信されたデータの内の音声データに基づいて音声を認識する音声認識手段と、前記音声認識手段により認識された音声に対応する文字を出力する文字出力手段と、前記受信手段により受信されたデータの内の複数の音声データに基づいて当該複数の音声データそれぞれの発生位置を検出する音声位置検出手段と、前記表示手段に表示されている顔画像の内、前記口元位置検出手段により検出された口元の変化領域の位置又はその近傍位置と前記音声位置検出手段により検出された各音声の発生位置に対応する表示位置との少なくとも一方の位置に、前記文字出力手段から出力された音声に対応する文字を表示制御する表示制御手段と、を有することを特徴とする。

【0012】以上の構成によれば、受信手段は外部から

送られてきた顔画像データ及び複数ヶ所より集音された複数の音声データを含むデータを受信し、表示手段は受信手段により受信されたデータの内の顔画像データに基づいて対応する顔画像を表示し、口元位置検出手段は受信手段により受信されたデータの内の顔画像データに基づいて口元の変化領域を検出し、音声認識手段は受信手段により受信されたデータの内の音声データに基づいて音声を認識し、文字出力手段は音声認識手段により認識された音声に対応する文字を出力し、音声位置検出手段は受信手段により受信されたデータの内の複数の音声データに基づいてこの複数の音声データそれぞれの発生位置を検出し、表示制御手段は、表示手段に表示されている顔画像の内、口元位置検出手段により検出された口元の変化領域の位置又はその近傍位置と音声位置検出手段により検出された各音声の発生位置に対応する表示位置との少なくとも一方の位置に、文字出力手段から出力された音声に対応する文字を表示制御する。

【0013】従って、受信された顔画像データに基づく顔画像を表示出力する際にその顔画像中に口元位置又はその近傍位置と音声の発生位置との少なくとも一方の位置に、受信された音声対応の文字の表示を制御するようにしたので、顔画像中の口元の動きが検出されない場合又は音声の受信が良好でなかった場合でも、受信された音声に対応する文字を顔画像の口元位置等に暫定的に表示出力することが可能となる。

【0014】請求項3記載の発明に係る顔画像表示装置は、請求項1又は2に記載の発明において、前記受信手段により受信されたデータの内の音声データに基づいて対応する音声を外部に出力する音声出力手段を更に有することを特徴とする。

【0015】以上の構成によれば、受信された音声対応の文字の表示とともにその音声を外部に出力するようにしたので、請求項1又は2に記載の発明の効果に加え、視覚と聴覚との相乗効果で音声内容を確認することが可能である。

【0016】請求項4記載の発明に係る顔画像表示装置は、請求項1又は2に記載の発明において、前記文字出力手段から出力される音声に対応する文字を包囲する吹出し画像が記憶されている吹出し記憶手段とを更に備えており、前記表示制御手段は前記文字出力手段から文字を出力した際に、その文字の表示領域を包囲する吹出し画像を前記吹出し記憶手段から読み出して合成表示する吹出し表示制御手段を含むことを特徴とする。

【0017】以上の構成によれば、受信された音声に対応する文字を表示の際にその文字を吹出し画像で包囲するようにしたので、請求項1又は2に記載の発明の効果に加え、文字を他の文字又は他の表示画像から判別しやすいように表示することが可能である。

【0018】請求項5記載の発明に係る顔画像表示装置は、請求項4記載の発明において、前記吹出し記憶手段

は、表示文字数に対応させてサイズの異なる複数の吹出し画像を記憶する吹出し画像記憶手段を備えており、前記表示制御手段は、前記文字出力手段から出力された音声に対応する文字の文字数に対応する吹出し画像を前記吹出し記憶手段に記憶された複数の吹出し画像の中から読み出して表示する読出表示手段を備えていることを特徴とする。

【0019】以上の構成によれば、音声認識された音声に対応する文字の数に応じたサイズの吹出し画像を採用するようにしたので、請求項4記載の発明の効果に加え、必要以上に吹出し画像の領域をとらず、表示画像の隠蔽を必要最小限に抑えることが可能である。

【0020】請求項6記載の発明に係る顔画像表示装置は、請求項4又は5に記載の発明において、さらに、前記受信手段により受信されたデータの内の顔画像データに基づいて前記表示手段の表示領域の内の特定領域を検出する特定領域検出手段を有し、前記吹出し表示制御手段は、前記受信手段により受信されたデータの内の顔画像データに基づく表示画面上で前記特定領域検出手段により検出された特定領域の表示領域とは別の領域に前記吹出し画像を表示することを特徴とする。

【0021】以上の構成によれば、特定領域を除いた表示領域に吹出し画像を表示するようにしたので、請求項4又は5に記載の発明の効果に加え、特定領域の表示画像を隠蔽せずに吹出し画像を表示することが可能である。

【0022】請求項7記載の発明に係る顔画像表示装置は、請求項3記載の発明において、前記文字出力手段から出力される音声に対応する文字を包囲する吹出し画像およびこの吹出し画像と前記口元位置検出手段により検出された口元位置とに配置される吹出し起点画像がそれぞれ記憶されている画像記憶手段とを更に備えており、前記表示制御手段は前記文字出力手段から文字を出力した際に、その文字の表示領域を包囲する吹出し画像と前記口元位置検出手段により検出された口元位置に配置される吹出し起点画像とを前記画像記憶手段から読み出して合成表示する吹出し表示手段を含むことを特徴とする。

【0023】以上の構成によれば、表示の際に音声に対応する文字を吹出し画像で包囲するようにしたので、請求項3記載の発明の効果に加え、文字を他の表示画像から判別しやすいように表示することが可能である。

【0024】請求項8記載の発明に係る顔画像通信システムは、請求項1記載の顔画像表示装置と、前記顔画像表示装置に対し回線を介してデータを送信する送信装置とからなる顔画像通信システムにおいて、前記送信装置は、顔画像データを入力する画像入力手段と、音声データを入力する音声入力手段と、前記画像入力手段により入力された顔画像データと前記音声入力手段により入力された音声データとをデータとして前記顔画像表示装置

10

20

30

40

50

に送信する送信手段と、を有することを特徴とする。

【0025】以上の構成によれば、送信装置において、画像入力手段は顔画像データを入力し、音声入力手段は音声データを入力し、送信手段は画像入力手段により入力された顔画像データと音声入力手段により入力された音声データとをデータとして顔画像表示装置に送信する。

【0026】従って、送信装置から顔画像表示装置に対して顔画像データと音声データとを送信するだけで、顔画像データに基づく顔画像を表示出力する際にその顔画像の口元の位置又はその近傍位置に、受信された音声対応の文字を同時に表示するようにしたので、簡単なシステム構成で音声に対応する文字と顔画像とを表示出力することが可能である。

【0027】請求項9記載の発明に係る顔画像通信システムは、請求項2記載の顔画像表示装置と、前記顔画像表示装置に対し回線を介してデータを送信する送信装置とからなる顔画像通信システムにおいて、前記送信装置は、複数の顔画像データを入力する画像入力手段と、複数ヶ所より集音された複数の音声データを入力する音声入力手段と、前記画像入力手段により入力された複数の顔画像データと前記音声入力手段により入力された複数の音声データとをデータとして前記顔画像表示装置に送信する送信手段と、を有することを特徴とする。

【0028】以上の構成によれば、送信装置において、画像入力手段は複数の顔画像データを入力し、音声入力手段は複数ヶ所より集音された複数の音声データを入力し、送信手段は、画像入力手段により入力された複数の顔画像データと音声入力手段により入力された複数の音声データとをデータとして顔画像表示装置に送信する。

【0029】従って、送信装置から顔画像表示装置に対して複数の顔画像データと複数の音声データとを送信するだけで、顔画像データに基づく顔画像を表示出力する際にその顔画像中に口元位置又はその近傍位置と音声の発生位置との少なくとも一方の位置に、受信された音声対応の文字の表示を制御するようにしたので、簡単なシステム構成で音声に対応する文字と顔画像とを表示出力することが可能である。

【0030】

【発明の実施の形態】以下に添付図面を参照して、本発明に係る好適な実施の形態を詳細に説明する。まず、本発明に係る顔画像通信システムについて説明する。図1は本発明に係る顔画像通信システムの一実施の形態を示す構成図である。この図1に示した顔画像通信システムにおいて、顔画像表示装置100と顔画像表示装置200とは、電話回線やRS232C等の回線LNによってデータ通信可能に接続されている。

【0031】顔画像表示装置100を送信側端末とした場合には、端末本体100aにCCDカメラISが接続され、これは撮影によって画像データを入力する。図1

の例では、そのCCDカメラISに対して例えば3人の人物（向かって左より甲、乙、丙）が対向して位置している。この3人（人物（甲）、（乙）、（丙））の人物画像は、CCDカメラISによって向かって左より人物画像SIMG1、SIMG2、SIMG3の並びでデータ入力される。これら人物画像SIMG1、SIMG2、SIMG3は後述する顔画像表示装置200に画像データとして送信されるものである。

【0032】また、端末本体100aには、CCDカメラISとともに2本のマイクML、MRが接続されており、一方のマイクMLは3人の人物に対して左側の音声を集音し、他方のマイクMRは3人の人物に対して右側の音声を集音する。この2本のマイクML、MRによってステレオ状態で音声が入力されることになり、人物（甲）が会話をした場合には、マイクMLから音声入力される音量がマイクMRから音声入力される音量よりも大きくなり（人物（甲）の向いている方向によっては異なる）、人物（乙）が会話をした場合には、マイクML及びマイクMRにそれぞれ音声入力される音量がほぼ等しくなり（人物（乙）の向いている方向によっては異なる）、人物（丙）が会話をした場合には、マイクMRから音声入力される音量がマイクMLから音声入力される音量よりも大きくなる（人物（丙）の向いている方向によっては異なる）。

【0033】このように、CCDカメラIS及びマイクML、MRからそれぞれ入力された画像データ（人物画像及び背景画像）及び音声データ（ステレオ音声）は、端末本体100aにより通信データとして回線LNに出力され、送信先である顔画像表示装置200に送信される。

【0034】顔画像表示装置200を受信側端末とした場合には、端末本体200aにCRTディスプレイDSPが接続され、これは表示画面30上に顔画像表示装置100から受信された画像データに基づく画像を表示する。図1の例に従い、そのディスプレイDSPには、3人の人物（甲）、（乙）、（丙）の人物画像RIMG1、RIMG2、RIMG3が左から右に向かって並んで表示されている。これら人物画像RIMG1、RIMG2、RIMG3は前述の顔画像表示装置100から画像データとして受信されるものである。

【0035】また、端末本体200aには、CRTディスプレイDSPとともに2台のスピーカSL、SRが接続されている。一方のスピーカSLは表示画面30に向かって左側に配置されており、前述のマイクSLより集音された音声を出力する。他方のスピーカSMは表示画面30に向かって右側に配置されており、前述のマイクMRにより集音された音声を出力する。この2台のスピーカSL、SRによってステレオ状態で音声が出力されることになり、人物（甲）が会話をした場合には、スピーカSLから音声出力される音量がスピーカSRから音

声出力される音量よりも大きくなり（人物（甲）の向いている方向によっては異なる）、人物（乙）が会話をした場合には、スピーカSL及びスピーカSRからそれぞれ音声出力される音量がほぼ等しくなり（人物（乙）の向いている方向によっては異なる）、人物（丙）が会話をした場合には、スピーカSRから音声出力される音量がスピーカSLから音声出力される音量よりも大きくなる（人物（丙）の向いている方向によっては異なる）。

【0036】このように、顔画像表示装置100より回線LNを介して通信データが顔画像表示装置200に受信されると、端末本体200aは、その通信データに基づいてCRTディスプレイDSP及び2台のスピーカSL、SRからそれぞれ人物画像（背景含む）及びステレオ音声を出力する。

【0037】例えば、人物（乙）がマイクML、MRに向かって「おはようございます」という会話SPKをした場合には、その会話による音声はマイクML、MRによって取り込まれて端末本体100aで音声データに変換される。この音声データはCCDカメラISで入力した画像データとともに通信データとなって顔画像表示装置200に送信される。顔画像表示装置200では、端末本体200aにおいて顔画像表示装置100から受信した通信データを画像データと音声データとに分け、その画像データに基づく画像はCRTディスプレイDSPによって表示出力され、同時に、その音声データに基づく音声SNDL、SNDL「おはようございます」はそれぞれ2台のスピーカSL、SRによって出力される。また、CRTディスプレイDSPでは、詳細については後述するが、その音声SNDL、SNDL「おはようございます」に対応する文字群WRD「おはようございます」が文字として表示出力される。

【0038】次に、顔画像表示装置100及び200について詳述する。なお、図1では送信側、受信側を説明する上で間略化して図示及びその説明を行っていたが、実際には顔画像表示装置100及び200は同一の構成を有していることから、以下に一方の顔画像表示装置200を例に挙げて説明する。

【0039】図2は顔画像表示装置200の内部構成の一例を示すブロック図、図3は図2に示した吹出しパターンROMのメモリ構成の一例を示す図、図4は図2に示した矢印パターンROMのメモリ構成の一例を示す図、図5、図8及び図10は図2に示した吹出しデータRAMのメモリ構成の一例を示す図、そして、図6は本実施の形態における画像合成手順を示す図である。

【0040】図2に示した顔画像表示装置200は、例えば、CPU2、ROM3、RAM4、キー入力部5、画像入力部6、音声入力部7、I/F8、送受信部9、音声認識部10、音量検出部11、表示駆動回路12、表示部13、音声発生部14、及びスピーカ部15を具備している。以上の構成において、端末本体200a

は、CPU2、ROM3、RAM4、送受信部9、音声認識部10、音量検出部11、及び表示駆動回路12を包含している。なお、顔画像表示装置100は、上記顔画像表示装置200と同様の構成を具備しており、以下にその顔画像表示装置100の動作を説明する場合には顔画像表示装置100の各部と同様の符号を用いるものとする。

【0041】まず、CPU2は装置全体を制御する中央処理部であり、ROM3に格納された各種制御プログラムに従って動作する。このCPU2は、図示の例では、ROM3、RAM4、キー入力部5、画像入力部6、音声入力部7、送受信部9、音声認識部10、音量検出部11、表示駆動回路12、及び音声発生部14の各ユニットに信号線で結合されており、これら各ユニットとアドレス信号、制御信号、データ等のやりとりからデータ入力（キー入力、音声認識データ、音量データ）、画像入力、音声入力、データ送受信、表示、プログラム実行、データのリード/ライト、音声発生等の各種制御を実行する。例えば、データ送信については、図12に示したフローチャートに従う制御プログラムによって制御が行われ、データ受信については、図13及び図14に従う制御プログラムによって制御が行われる。

【0042】ROM3は吹出しパターンROM31、矢印パターンROM32、辞書ROM33、CPU2を動作させるための制御プログラム、及び各種パラメータを記憶したメモリであり、CPU2の制御により各制御プログラムの読み出し、後述の各吹出しパターンデータ、矢印パターンデータ、辞書データが読み出される。

【0043】このROM3において、吹出しパターンROM31には、図3に示したように、一例としてパーツNo. P01～P20まで20種類の吹出しパターンデータが後述する認識文字数に対応させて記憶されている。この吹出しパターンデータに基づいて形成される吹出し画像である吹出しパターンは、一文節中にある文字数に対応させて $n \times m$ ドットのマトリックスで構成されるものである。この一文節中にある文字数は、後述する音声認識部10において一文節中から得られた認識文字数に相当する。例えば、認識文字数が3個であるという認識結果が得られた場合には、その3個に該当するパーツNo. が検索され、その結果検出されたパーツNo. P03に基づきそのパーツNo. P03の $N3 \times M3$ ドットの吹出しパターンが選択される。

【0044】矢印パターンROM32には、図4に示したように、一例としてパーツNo. D01～D20まで20種類の矢印パターンデータが吹出しパターンを吹出す方向すなわち吹出し方向に対応させて記憶されている。この矢印パターンデータに基づいて形成される矢印パターンは、前述の吹出しパターンの吹出し起点位置（すなわち、表示中の顔画像の口元の近傍位置）に配置される吹出し起点画像であり、吹出し方向が決定する

と、後述する吹出し画像RAM4 2において上述の吹出しパターンに合成される。例えば、認識文字数が3個であるという認識結果が得られ、かつその文字群を吹出しパターンによって口元位置より右上に方向に吹出す場合には、その認識文字数3個に対応するN3×M3ドットの吹出しパターンに、吹出し方向を右上とするパーツNo. D01の矢印パターンが選択される。

【0045】辞書ROM33は、音声認識部10によって得られた認識結果から文字パターンを発生するためのキャラクタジェネレータ機能を具備している。

【0046】RAM4は、吹出しデータRAM4 1、吹出し画像RAM4 2、合成RAM4 3、及びCPU2のワークRAMを具備している。

【0047】吹出しデータRAM4 1は、図5や図10に示したように、音声毎に吹出しデータNo. をNo. 1、No. 2…のように付してエリア分けを行い、各吹出しデータNo. に対応させて、口元位置データ（座標データ）、吹出しパターン（パーツNo.）及びその起点位置（座標データ）、矢印パターン（パーツNo.）及びその起点位置（座標データ）、文字群データ（文字コード）及びその起点位置（座標データ）を格納する。また、この吹出しデータRAM4 1では、後述するが、図8に示した如く、連続する音声を文節によって区切った場合に、吹出しデータNo. をNo. 1-1、No. 1-2…のように関連性をもたせて格納しており、その内容は図や図10の場合と同様である。

【0048】吹出し画像RAM4 2は、受信された画像データ上で人物画像（特定領域）を除いた領域に吹出しパターン及びその吹出しパターン内の文字パターンを合成するメモリである。合成RAM4 3は、受信された画像データに基づいて画像をドット展開して、吹出し画像RAM4 2でつくられた特定領域を除く領域のドットデータを合成するメモリである。この合成RAM4 3にドット展開された画像データは、表示部13において画像を表示するための表示データとして使用される。

【0049】キー入力部5は、文字、ファンクションキー等のキーを具備しており、CPU2の動作にかかるキー操作を必要としているときに使用される。この実施の形態では、例えば、送受信の際にキー入力部5の操作が必要となる。

【0050】画像入力部6は図1に示したCCDカメラISを有する撮影用ユニットであり、音声入力部7は図1に示した2本のマイクML、MRを有するステレオ音声入力用ユニットである。I/F8は回線LNに接続されるインタフェースであり、送受信部9はI/F8とCPU2とに接続され、通信データの送受信を行うものである。

【0051】音声認識部10は送受信部9とCPU2とに接続され、受信された通信データから音声データを取り出して音声認識を行ってその認識結果（認識データ）

をCPU2に供給するものである。音量検出部11は送受信部9とCPU2とに接続され、受信された通信データから音声データを取り出して音量を検出し、その検出結果（音量データ）をCPU2に供給するものである。

【0052】表示駆動回路12はCPU2の制御により表示部13の表示駆動を行う回路であり、表示部13は表示駆動回路12の表示駆動により表示データに基づく表示画像を形成する。音声発生部14はCPU2の制御により音声データに基づく音声を発生するものであり、スピーカ部15は音声発生部14に接続され、音声発生部14により発生した音声を外部に放音する2台のスピーカSL、SRを有する音声出力用ユニットである。

【0053】次に、動作について説明する。図7、図9及び図11は本実施の形態における送受信画像の対応関係の一例を示す図である。図12は本実施の形態による送信時の動作を説明するフローチャートであり、図13及び図14は本実施の形態による受信時の動作を説明するフローチャートである。なお、送信端末、受信端末はそれぞれ顔画像表示装置100、顔画像表示装置200として説明する。

【0054】まず、送信動作について図1、図7、図9、図11及び図12を参照して説明する。以下に説明する動作は、顔画像表示装置100のCPU2により制御されるものであり、個々の動作は各部で行われる。

【0055】図1に示した顔画像表示装置100では、キー入力部5による送信操作が検出されると、画像入力部6（CCDカメラIS）及び音声入力部7（マイクML、MR）を駆動して、3人（人物（甲）、（乙）、（丙））の画像及びその音声を取り込む処理が開始される（ステップS1）。このように、画像及び音声の入力が開始されると、図7（a）、（b）及び（c）、図9（a）及び（b）、図11（a）、（b）に示した画像及び音声は通信データに変換されて送信先の顔画像表示装置200に回線LNを介して送信される。この場合、通信データにおいては、回線LNの伝送速度等の伝送条件に応じてその情報量が決定される。

【0056】図7に示した例では、人物（甲）、（乙）、（丙）がその並びで画像入力部6に対面しており、これを画像入力部6が撮影して画像SIMG1、SIMG2、SIMG3を取り込むことにより、図7（a）に示した如く全体の画像データIMG1がRAM4の図示せぬ入力画像用のメモリに格納される。この場合には、3人の内のいずれからも音声が発せられていないが、音声入力部7のマイクML、MRによる2つの集音があることから音声データは存在することになる。

【0057】ステップS2では、入力された画像データIMG1及び音声データ（ステレオ状態）が通信データに変換され、送受信部9により変調されてI/F8より回線LNに送出される。すなわち、画像及び音声の送信が実行される。この画像及び音声の入力/送信動作は、

キー入力部5において終了操作が行われるまで繰り返し実行される(ステップS3)。その際、顔画像表示装置200には、回線LNを介して通信データが受信され、図13及び図14に示したフローに従う動作が開始される。

【0058】顔画像表示装置100では、継続して音声入力部7により音声を取り込まれ、画像入力部6により画像データIMG2が取り込まれ(図7(b)参照)、3人の中央に位置する人物(乙)が画像SIMG2に示したように音声「おはよう!」の挨拶をすると、音声入力部7には「おはよう!」という音声を取り込まれこれがCPU2により音声データに変換される。この後、この音声データは画像データIMG2とともに通信データとして顔画像表示装置200に送信される。

【0059】さらに継続して、音声入力部7により音声を取り込まれ、画像入力部6により画像データIMG3が取り込まれると(図7(c)参照)、向かって3人の一番左側に位置する人物(甲)が画像SIMG1に示したように音声「お元気ですか?」の挨拶をすると、音声入力部7には「お元気ですか!」という音声を取り込まれこれがCPU2により音声データに変換される。この後、この音声データは画像データIMG3とともに通信データとして顔画像表示装置200に送信される。

【0060】このように、送信側の顔画像表示装置100は、送信終了の指示があるまで、順次画像及び音声を取り込んでこれを通信データとして受信側の顔画像表示装置200に送信する動作を繰り返す。なお、図9及び図11についても同様の手順で画像及び音声送信が実行される。

【0061】図9(a)及び(b)の例では、画像データIMG11(図9(a))から画像データIMG12(図9(b))への移行にみられるように、人物(乙)が音声「おはようございます! お元気ですか?」を発しており、これは送信の段階ではひとつの音声データ(ステレオ状態)として送信されるが、受け手の顔画像表示装置200では音声認識後に文字群が文節で区切られることから、文字群表示の際に、第1文字群「おはようございます!」と第2文字群「お元気ですか?」との2文となる。

【0062】図11(a)及び(b)の例では、画像データIMG21(図11(a))から画像データIMG22(図11(b))への移行にみられるように、音声入力部7より音声「おはようございます!」が取り込まれているにもかかわらず、画像データIMG22からは口元の動きが読みとれず3人の内の誰が音声を発しているのか不明となる。このため、この実施の形態では、マイクML、MRを一定の間隔を置いて配置することを前提に、マイクMLとMRとを図1の例では3人の両端

(人物(甲)側と人物(丙)側)にそれぞれ配置させることを条件とすれば、顔画像表示装置200において音

声の発生元を2点(マイクML、MR)の検出音量の差異から暫定的に決定することが可能となる。

【0063】次に、受信動作について図1並びに図5～図14を参照して説明する。以下に説明する動作は、顔画像表示装置200のCPU2により制御されるものであり、個々の動作は各部で行われる。

【0064】図1に示した顔画像表示装置200では、顔画像表示装置200との通信準備が整うと、ステップS11において、前述の図2のフローチャートに従って送信されてくる通信データを受信する動作が開始される。

【0065】ステップS11において画像及び音声受信が開始されると、ステップS12において、送受信部9で受信された通信データの内の音声データに基づく音声認識とその音量検出とが開始され、これらの処理は音声認識部10、音量検出部11で行われる。

【0066】ステップS13では、送受信部9で受信された通信データの内の画像データは合成RAM43にドット展開され、表示部13に受信画像が表示されるとともに、音声発生部14では、受信された音声データに基づいて音声が発生され、その音声スピーカ部15のスピーカSL、SRよりステレオ状態で放音される。すでに送信動作で説明した図7(a)、(b)、(c)の例を挙げると、顔画像表示装置100では、図7(b)の段階で人物(乙)が音声「おはよう!」を発している人物画像SIMG2及びその音声を取り込む処理が実行され、一方、これを通信データとして受け取った顔画像表示装置200では、表示部13の表示画面30に受信画像を表示するとともに、左右のスピーカSL、SRから均等な音量で音声認識された音声「おはよう!」を出力する処理が実行される。

【0067】なお、人物(乙)がマイクML、MR間のほぼ中央に位置しているため、スピーカSL、SRの出力音量は均等となる。この受信側での画像表示と音声出力とが行われる際にステップS13以降の処理が実行される。

【0068】次のステップS14では、図7(b)に示した画像データIMG2が受信された場合、その画像データIMG2はI IMG2として吹出し画像RAM42にドット展開される。この画像データI IMG2はステップS13においてすでに合成RAM43にも格納されている。続くステップS15及びS16において、図6(a)に示したように吹出し画像RAM42にドット展開された画像データI IMG2から特定領域すなわち人物画像SIMG1、2、3に対応する人物領域SS IMG1、2、3の各画像データが一定の色属性をもつドットにマスク(図中、斜線で示す)される。その結果、吹出し画像RAM42では、このマスクされた人物領域SS IMG1、2、3を除く領域が背景領域BKとして設定され

る。

【0069】なお、ステップS15における人物領域SSIMG1, 2, 3の検出には、使用者がキー入力部5を操作して表示部13の表示画像に対して設定する等の方法を適用することができる。

【0070】次に、ステップS17では、上述したステップS15において検出された人物領域SSIMG1, 2, 3に対応する合成RAM43の画像データIIMG2から所定の変化を行う領域を検出する処理が開始される。この所定の変化を行う領域とは、口の開閉動作（口 10 元の動き）に相当する口元領域をいう。

【0071】図6(c)に示した合成RAM43において、その画像データIIMG2から口元領域32AAが検出されると（ステップS18）、その口元領域32AAに対応する口元画像MVが図6(b)の画像データIIMG2より決定され、その口元画像MVの座標位置が口元座標データとして吹出しデータRAM41に格納される（ステップS19）。その後、処理はステップS20（図14参照）に移行する。図7(b)に示した画像 20 IIMG2及びその音声に対応する画像データIIMG2及びその音声データにかかるデータを吹出しデータNo. 1とした場合、図5に示した如く、口元位置データはMD1として記憶される。

【0072】なお、図7(a)に示した画像データIMG1及びその音声データが受信された場合には、まず口元の動きが検出できないことから、ステップS18において口元の動きなしという判定がなされ、処理がステップS21（図14参照）に移行する。このステップS21では、その受信された画像データIMG1に対応する 30 音声データから音声認識が行われ、音声認識ができなかった場合には処理はステップS30に移行し、一方、音声認識ができた場合には処理は次のステップS22に移行する。

【0073】ここで処理の説明をステップS20に戻し、このステップS20では、受信された画像データIMG1に対応する音声データから音声認識が行われ、音声認識ができなかった場合には処理はステップS30に移行し、一方、音声認識ができた場合には処理は次のステップS24に移行する。図7(b)の例では、音声認識部10より音声認識できた音声データがCPU2に供給されるので、CPU2は辞書ROM33を参照してその音声データに基づく文字群データを発生させる。その際、ステップS24において、その文字群データから文 40 節分けが必要か否かの判断がなされる。図7(b)の例では人物(乙)が音声「おはよう！」の一文節のみであることから、ステップS24において文節分け不要という判断がなされ、ステップS26において、図5に示した如く文字群データがCHR1として吹出しデータRAM41に格納される。なお、ステップS24において文節分け必要という判断がなされた場合には、後述する 50

が、図8に示した如く、吹出しデータNo. 1、No. 2…のように文節分けした数に応じた吹出しデータの格納領域が設定される。

【0074】続くステップS27では、上述した口元位置データMD1、文字群データCHR1、及び背景領域BK（図6(b)参照）の座標位置データに基づいて、表示画面30に表示させる際の文字群のサイズ（文字列の段数や文字列の文字数など）、吹出し位置及び吹出し方向を算出して求め、その算出結果に対応する吹出しパターン及びその矢印パターンと、これらパターンの各起 10 点位置（配置させる基準の位置）とを決定する処理が実行される。

【0075】上述した図7(b)の例では、文字群データCHR1に対応する文字群は「おはよう」となっているので、その文字数は4文字である。この4文字の文字群「おはよう」については、ステップS27において、一列で口元画像MV近傍の背景領域BKに配置できるという結果からその起点位置(XC, YC)が得られ、かつ、その一列4文字を包囲できる吹出しパターンとして認識文字数4文字に対応するパーツNo. P04の吹出しパターン及びその起点位置(XA, YA)が決定される。また、吹出し方向は、図6(b)に示したように右上との決定が得られたことから、矢印パターンはパーツ 20 No. D01がその起点位置(XB, YB)とともに決定される。

【0076】このように、文字群及びその起点位置、吹出しパターン及びその起点位置、ならびに、矢印パターン及びその起点位置が決定されると、これらの各データCHR1、(XC, YC)、P04、(XA, YA)、D01、(XB, YB)が吹出しデータNo. 1の所定の格納エリア（吹出しデータRAM41）に格納される。

【0077】続くステップS28では、吹出しデータRAM41に格納された各データに基づいて図6(b)に示した如く吹出し画像RAM42の背景領域BKに画像データがドット展開される。すなわち、吹出し画像RAM42には、吹出しパターンデータに基づく吹出しパターンP04がドット展開され、この吹出しパターンP04に対して矢印パターンデータに基づく矢印パターンD01が合成してドット展開され、さらに文字群データCHR1に基づく文字パターンが吹出しパターンP04内に収まるようにドット展開される。このようにして、吹出しパターンP04、矢印パターンD01、及び文字群「おはよう」の各データよりなる吹出し画像データ30AAが形成される。

【0078】次のステップS29では、合成RAM43にすでに展開されている画像データIIMG2に対して、上記ステップS28により吹出し画像RAM42に展開された背景領域BK部分の画像データが合成される。その結果、図6(c)に示した如く、口元領域32 50

AA近傍に上述した吹出し画像データ30AAが合成されるので、表示部13の表示画面30には、図7(b)に示したように、人物(乙)の人物画像RIMG2の口元32Aから吹出すように吹出し画像30A「おはよう」が表示され、その際に、左右のスピーカSL, SRよりほぼ均等の音量で音声「おはよう」が出力される。受信側では、吹出し画像30A中の吹出し位置31Aからだれからの吹き出しであるのかを口元の動きとともに確認することができる。

【0079】そして、キー入力部5の操作で終了が指示されるまでは、処理は再びステップS13に戻り、上述したステップS13～S29までの動作を繰り返し実行する(ステップS30)。

【0080】続いて、図7(b)の状態から同図(c)の状態に送信画像が推移した場合には、向かって一番左に位置する人物(甲)(人物画像SIMG1参照)が音声「お元気ですか?」を発していることから、人物

(乙)の場合と同様に、表示部13の表示画面30には、人物(甲)の人物画像RIMG1の口元32Bから吹出すように吹出し画像30B「おげんきですか」(この場合には、文字群の文字数と背景領域のサイズとの関係から文字列が2段となる)が表示され、その際に、スピーカSLがスピーカSRよりも音量が大きくなるように音声「お元気ですか?」が出力される。人物(甲)はマイクML近傍に位置しており、もうひとつのマイクMRからは離れているので、マイクMLによって集音される音量は高くなるが、これに比べてマイクMRによって集音される音量は低くなるので、スピーカSL, SRがこれに対応してステレオ状態で音声を出力する。この集音された音量の違いは、受信側において音量検出部11によって検出され、後述するが、口元の動きが検出できなかったときに暫定的に音声発生位置を決定するために採用される。このように、受信側では、吹出し画像30B中の吹出し位置31Bからだれからの吹き出しであるのかを口元の動きとともに確認することができる。

【0081】なお、吹出しデータRAM41には、今度は吹出しデータNo. 2として、口元位置データMD2、吹出しパターンP07その起点位置(XD, YD)、矢印パターンD01及びその起点位置(XE, YE)、文字群データCHR2及びその起点位置(XF, YF)が記憶される。

【0082】また、図9(a)及び(b)の各画像データIMG11, IMG12に示したように、送信側(顔画像表示装置100)でひとりの人物(乙)(人物画像SIMG2)により発せられた音声「おはようございます! お元気ですか?」から、図9(a)、(b)及び(c)の各表示画面30に示したように、受信側(顔画像表示装置200)で複数の文節(第1文字群「おはようございます」と第2文字群「おげんきですか」)に分けて各文節に対応した表示画像(図9(b)及び

(c))を連続して表示する必要があると、前述のステップS24(図14参照)において、文節分けが必要であると判断され、処理はステップS25に移行する。

【0083】このステップS25では、図8に示した如く、文節分けにより2つの吹出しデータは関連性あるものとなり、予め吹出しデータRAM41には、吹出しデータNo. 1及びNo. 2の格納領域が確保される。その際、各吹出しデータの格納領域において、口元位置データ(例えばMD3)は共通化して格納される。

【0084】続くステップS26～S29までは前述と同様に処理が実行されるが、その際、各ステップにおいて吹出しデータNo. 1とNo. 2とにそれぞれペアでデータを格納する処理が行われる。

【0085】吹出しデータRAM41において、吹出しデータNo. 1には、口元位置データMD3、第1文字群の9文字に対応する吹出しパターンP09その起点位置(XG, YG)、吹出しパターンP09の吹出し方向が中上であることを示す矢印パターンD02及びその起点位置(XE, YE)、第1文字群「おはようございます」の文字群データCHR3及びその起点位置(XI, YI)が記憶される。また、吹出しデータNo. 2には、吹出しデータNo. 1と同様の口元位置データMD3、第2文字群の7文字に対応する吹出しパターンP07その起点位置(XJ, YJ)、吹出しパターンP07の吹出し方向が中上であることを示す矢印パターンD02及びその起点位置(XK, YK)、第2文字群「おげんきですか」の文字群データCHR4及びその起点位置(XL, YL)が記憶される。

【0086】表示部13の表示画面30には、第1文節すなわち第1文字群について(図9(b)参照)、人物(乙)の人物画像RIMG2の口元32C近傍から吹出すように吹出し画像30C「おはようございます」が表示され、その際に、左右のスピーカSL, SRよりほぼ均等の音量で音声「おはようございます」が出力される。受信側においては、吹出し画像30C中の吹出し位置31Cからだれからの吹き出しであるのかを口元の動きとともに確認することができる。続く第2文節すなわち第2文字群について(図9(c)参照)、人物(乙)の人物画像RIMG2の口元32D近傍から吹出すように吹出し画像30D「おげんきですか」が表示され、その際に、左右のスピーカSL, SRよりほぼ均等の音量で音声「お元気ですか?」が出力される。受信側では、吹出し画像30D中の吹出し位置31Dからだれからの吹き出しであるのかを口元の動きとともに確認することができる。

【0087】この図9に示した表示画面30中の口元(31C, 31D)と吹出し画像(30C, 30D)との位置関係は、第1文字群、第2文字群のいずれも図7に示した表示画面30中の口元31Aと吹出し画像30Aとの位置関係よりも離間しており、これは、各文字群

の文字数を包囲する吹出しパターンのサイズとその吹出しパターンを配置できる背景領域との関係から決定されるものである。

【0088】また、図11(a)及び(b)の各画像データIMG21、IMG22に示したように、送信側(顔画像表示装置100)で口元の動きが無くひとりの人物により発せられた音声「おはようございます!」があると、図11(a)及び(b)の各表示画面30に示したように、受信側(顔画像表示装置200)では暫定的に吹出し画像30Eを表示して、これに合せてスピーカ部15より音声出力する処理が行われる。

【0089】具体的には、前述のステップS21(図14参照)において音声認識ができたという判定がなされると、処理はステップS22に移行し、音量検出部11にて検出されたマイクMLに対応する左側音量データとマイクMRに対応する右側音量データとに基づいて左右の音量差が算出される。続くステップS23において、その音量差から口元位置データが暫定的に求められ、その口元位置データは図10に示した如く暫定フラグFとともにMD4として吹出しデータNo. 1(吹出しデータRAM41)の格納領域に格納される。

【0090】その後、処理はステップS24に移行し、この場合には文節分けが不要なことから、さらにステップS26以降の処理が前述した通り実行される。

【0091】図10に示した吹出しデータNo. 1には、暫定的に決定された口元位置データMD4及び暫定フラグF、文字群の9文字に対応する吹出しパターンP09及びその起点位置(XM, YM)、吹出しパターンP09の吹出し方向が暫定的に中上であることを示す矢印パターンD02及びその起点位置(XN, YN)、文字群「おはようございます」の文字群データCHR5及びその起点位置(XP, YP)が記憶される。

【0092】表示部13の表示画面30には(図11(b)参照)、暫定的に人物(乙)の人物画像RIMG2の口元近傍から吹出すように吹出し画像30E「おはようございます」が表示され、その際に、左右のスピーカSL, SRよりほぼ均等の音量で音声「おはようございます」が出力される。受信側においては、吹出し画像30E中の吹出し位置31Eからどれからの吹き出しであるのかを口元の動きが無くても確認することができる。その際、スピーカSL, SRから出力される音量差の助けも借りて音声発生源を容易に推測することができる。

【0093】以上説明したように本実施の形態によれば、顔画像表示装置間で画像データと音声データとを送信し、受信側において画像データに基づく人物画像と音声データに基づく会話の音声とを出力する際にその人物画像の口元近傍にその音声対応の文字群を同時に表示するようにしたので、画像及び音声を送信する簡単なシステムにおいてその音声に対応する文字群を人物画像や音

声とともに口元を音声発生位置にみだてて同時に表示出力することが可能である。その結果、複数の顔画像から同時に音声対応の文字が表示された場合でも、その各文字と複数の顔画像との対応関係を明瞭に視認することができる。

【0094】また、顔画像を表示出力する際に、その顔画像中に口元位置又はその近傍位置と音声の発生位置との少なくとも一方の位置に、受信された音声対応の文字を表示制御するようにしたので、顔画像中の口元の動きが検出されない場合又は音声の受信が良好でなかった場合でも、受信された音声に対応する文字を顔画像の口元位置等に暫定的に表示出力することが可能である。

【0095】また、表示の際に文字群を吹出しパターンで包囲するようにしたので、文字群を他の表示画像から判別しやすいように表示することが可能である。

【0096】また、音声認識された文字の数に応じたサイズの吹出しパターンを採用するようにしたので、必要以上に吹出しパターンの表示領域をとらずに済み、これによって、他の表示画像の隠蔽を必要最小限に抑えることが可能である。

【0097】また、人物画像領域等の特定領域を除いた表示領域に吹出しパターンを合成表示するようにしたので、その特定領域の画像を隠蔽せずに吹出しパターンを表示することが可能となり、これによって、遠隔地の相手の表情や動きを十分に読み取りながら、提示された資料を見逃すことのないように、及び伝達情報に欠落のないように、会話を進めることができる。

【0098】さて、前述の実施の形態では、吹出しパターンの輪郭形状は図3に示したように矩形となっているが、矩形以外の楕円、円、星、雲、多角形等であってもよく、あるいは、これら各輪郭形状を予め吹出しパターンROM31に記憶しておき、キー入力部5の操作で選択できるようにしてもよく、吹出しパターンの選択幅を広げて使用者の用途や好みに応じた吹出しパターンを適宜提供することが可能である。

【0099】また、前述の実施の形態では、吹出しパターンと矢印パターンとをそれぞれ独立して記憶していたが、吹出しパターンと矢印パターンとを一体化したパターンを一元管理する形態にしてもよく、これによって吹出しパターンと矢印パターンとを合成する処理を省くことができる。

【0100】また、前述の実施の形態では、人物画像の表示領域を特定領域としていたが、人物画像の顔部分を特定領域としてもよく、この場合にも少なくとも誰が音声を発しているのかを判別することに支障はない。

【0101】また、前述の実施の形態において、特定領域を判別する際に、キー入力部5のキー操作で設定する方法以外に、人物の背景に所定の色(例えば青色)で印刷されたスクリーン等を配置し、受信側で画像処理によってその色を検出することで、人物画像と特定領域とを

境界を判別できるようにしてもよい。

【0102】また、前述の実施の形態では、文字群の表示単位を文節としていたが、その文節以外に、単語や文章を表示単位としてもよい。

【0103】

【発明の効果】以上説明したように請求項1記載の発明によれば、受信された顔画像データに基づく顔画像を表示出力する際にその顔画像の口元の位置又はその近傍位置に、受信された音声対応の文字を同時に表示するようにしたので、複数の顔画像から同時に音声対応の文字が表示された場合でも、その各文字と複数の顔画像との対応関係を明確に視認することが可能な顔画像表示装置を得られるという効果を奏する。

【0104】請求項2記載の発明によれば、受信された顔画像データに基づく顔画像を表示出力する際にその顔画像中に口元位置又はその近傍位置と音声の発生位置との少なくとも一方の位置に、受信された音声対応の文字の表示を制御するようにしたので、顔画像中の口元の動きが検出されない場合又は音声の受信が良好でなかった場合でも、受信された音声に対応する文字を顔画像の口元位置等に暫定的に表示出力することが可能な顔画像表示装置を得られるという効果を奏する。

【0105】請求項3記載の発明によれば、受信された音声対応の文字の表示とともにその音声を外部に出力するようにしたので、請求項1又は2に記載の発明の効果に加え、視覚と聴覚との相乗効果で音声内容を確認することが可能な顔画像表示装置を得られるという効果を奏する。

【0106】請求項4記載の発明によれば、受信された音声に対応する文字を表示の際にその文字を吹出し画像で包囲するようにしたので、請求項1又は2に記載の発明の効果に加え、文字を他の文字又は他の表示画像から判別しやすいように表示することが可能な顔画像表示装置を得られるという効果を奏する。

【0107】請求項5記載の発明によれば、音声認識された音声に対応する文字の数に応じたサイズの吹出し画像を採用するようにしたので、請求項4記載の発明の効果に加え、必要以上に吹出し画像の領域をとらず、表示画像の隠蔽を必要最小限に抑えることが可能な顔画像表示装置を得られるという効果を奏する。

【0108】請求項6記載の発明によれば、特定領域を除いた表示領域に吹出し画像を表示するようにしたので、請求項4又は5に記載の発明の効果に加え、特定領域の表示画像を隠蔽せずに吹出し画像を表示することが可能な顔画像表示装置を得られるという効果を奏する。

【0109】請求項7記載の発明によれば、表示の際に音声に対応する文字を吹出し画像で包囲するようにしたので、請求項3記載の発明の効果に加え、文字を他の表示画像から判別しやすいように表示することが可能な顔画像表示装置を得られるという効果を奏する。

【0110】請求項8記載の発明によれば、送信装置から顔画像表示装置に対して顔画像データと音声データとを送信するだけで、顔画像データに基づく顔画像を表示出力する際にその顔画像の口元の位置又はその近傍位置に、受信された音声対応の文字を同時に表示するようにしたので、簡単なシステム構成で音声に対応する文字と顔画像とを表示出力することが可能な顔画像通信システムを得られるという効果を奏する。

【0111】請求項9記載の発明によれば、送信装置から顔画像表示装置に対して複数の顔画像データと複数の音声データとを送信するだけで、顔画像データに基づく顔画像を表示出力する際にその顔画像中に口元位置又はその近傍位置と音声の発生位置との少なくとも一方の位置に、受信された音声対応の文字の表示を制御するようにしたので、簡単なシステム構成で音声に対応する文字と顔画像とを表示出力することが可能な顔画像通信システムを得られるという効果を奏する。

【図面の簡単な説明】

【図1】本発明に係る顔画像通信システムの一実施の形態を示す構成図である。

【図2】本発明に係る顔画像表示装置の一実施の形態を示すブロック図である。

【図3】図2に示した吹出しパターンROMのメモリ構成の一例を示す図である。

【図4】図2に示した矢印パターンROMのメモリ構成の一例を示す図である。

【図5】図2に示した吹出しデータRAMのメモリ構成の一例を示す図である。

【図6】本実施の形態における画像合成手順を示す図である。

【図7】本実施の形態における送受信画像の対応関係の一例を示す図である。

【図8】図2に示した吹出しデータRAMのメモリ構成の他の例を示す図である。

【図9】本実施の形態における送受信画像の対応関係の他の例を示す図である。

【図10】図2に示した吹出しデータRAMのメモリ構成の他の例を示す図である。

【図11】本実施の形態における送受信画像の対応関係の他の例を示す図である。

【図12】本実施の形態による送信時の動作を説明するフローチャートである。

【図13】本実施の形態による受信時の動作（前半）を説明するフローチャートである。

【図14】本実施の形態による受信時の動作（後半）を説明するフローチャートである。

【符号の説明】

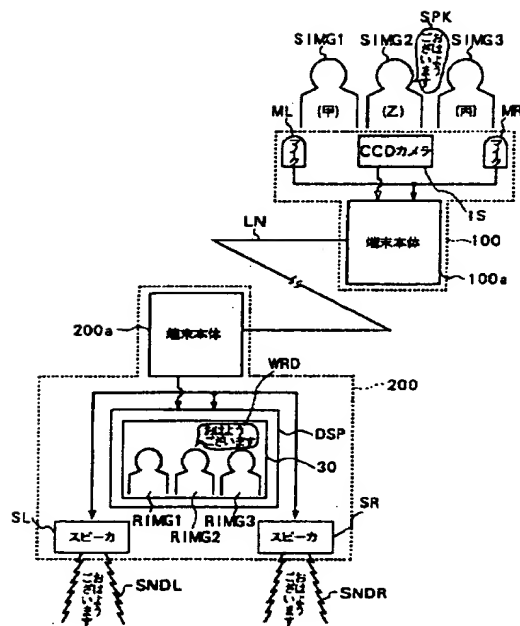
100, 200 顔画像表示装置

100a, 200a 端末本体

2 CPU

- 3 ROM
- 4 RAM
- 5 キー入力部
- 6 画像入力部
- 7 音声入力部
- 8 I/F
- 9 送受信部
- 10 音声認識部
- 11 音量検出部
- 12 表示駆動回路

【図1】



【図4】

パーツNo.	D01	D02	D03	D20
吹出し方向	右上	中上	左上	右下
矢印パターン				

- 13 表示部
- 14 音声発生部
- 15 スピーカ部
- 30 表示画面
- 31 吹出しパターンROM
- 32 矢印パターンROM
- 33 辞書ROM
- 41 吹出しデータRAM
- 42 吹出し画像RAM
- 10 43 合成RAM

【図3】

パーツNo.	P01	P02	P03	P20
認識文字数 (画)	1	2	3	20
吹出しパターン (nXmドット)				
	N1XM1 (ドット)	N2XM2 (ドット)	N3XM3 (ドット)	N20XM20 (ドット)

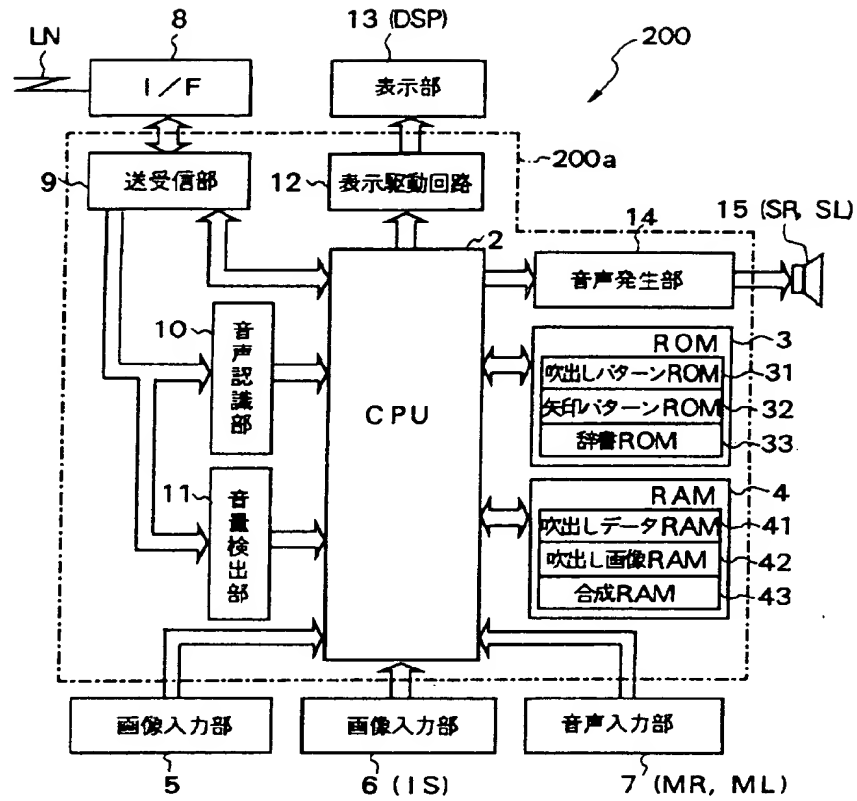
【図5】

【図10】

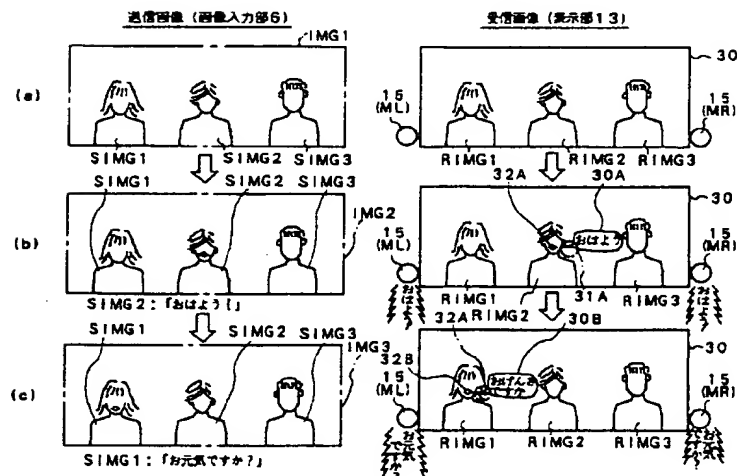
吹出しデータNo. 1	
口元位置データ	MD1
吹出しパターン	P04
起点位置	XA, YA
矢印パターン	D01
起点位置	XB, YB
文字群データ	CHR1
起点位置	XC, YC
吹出しデータNo. 2	
口元位置データ	MD2
吹出しパターン	P07
起点位置	XD, YD
矢印パターン	D01
起点位置	XE, YE
文字群データ	CHR2
起点位置	XF, YF

吹出しデータNo. 1	
口元位置データ	MD4 (F)
吹出しパターン	P09
起点位置	XM, YM
矢印パターン	D02
起点位置	XN, YN
文字群データ	CHR5
起点位置	XP, YP

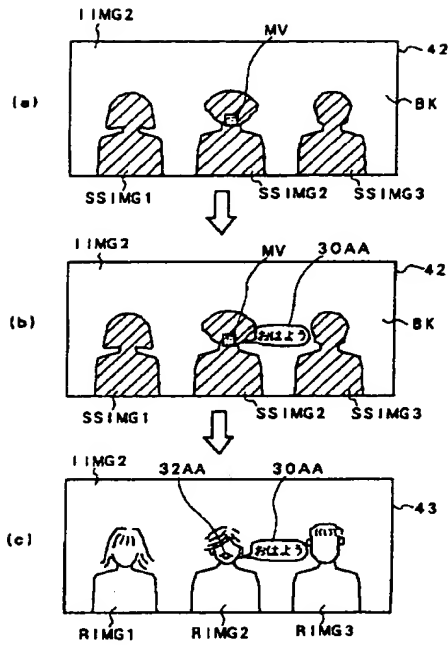
【図2】



【図7】



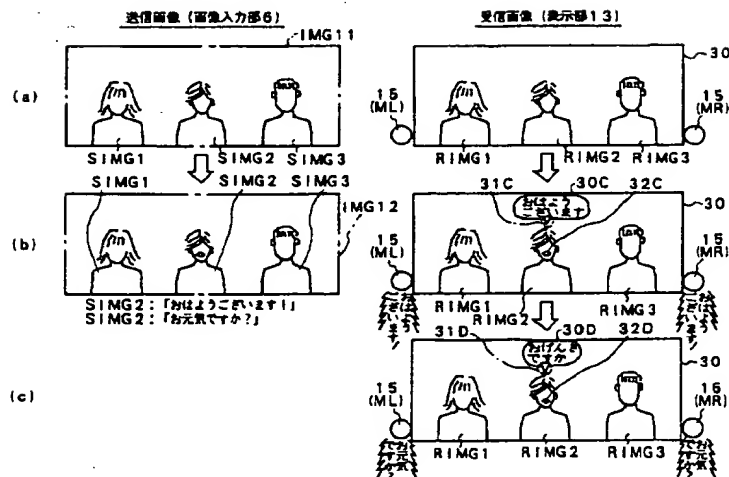
【図6】



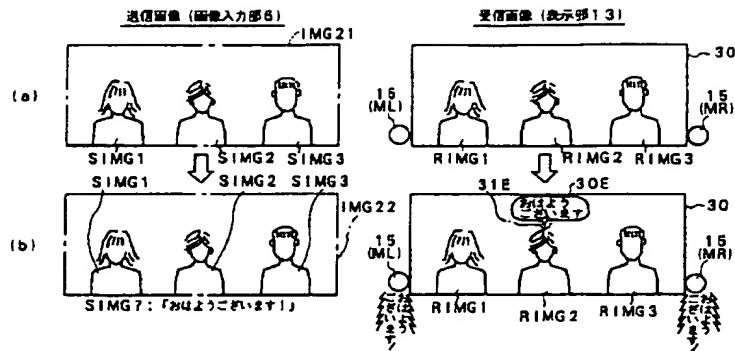
【図8】

送出しデータNo.1		41
口元位置データ	MD3	
吹き出しパターン	P08	
発点位置	XG, YG	
矢印パターン	D02	
終点位置	XH, YH	
文字群データ	CHR3	
発点位置	XI, YI	
送出しデータNo.2		
口元位置データ	MD3	
吹き出しパターン	P07	
発点位置	XJ, YJ	
矢印パターン	D02	
終点位置	XK, YK	
文字群データ	CHR4	
発点位置	XL, YL	

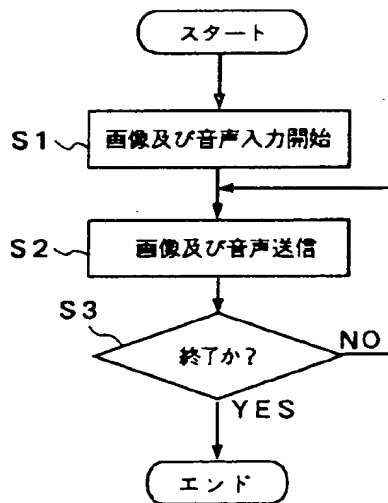
【図9】



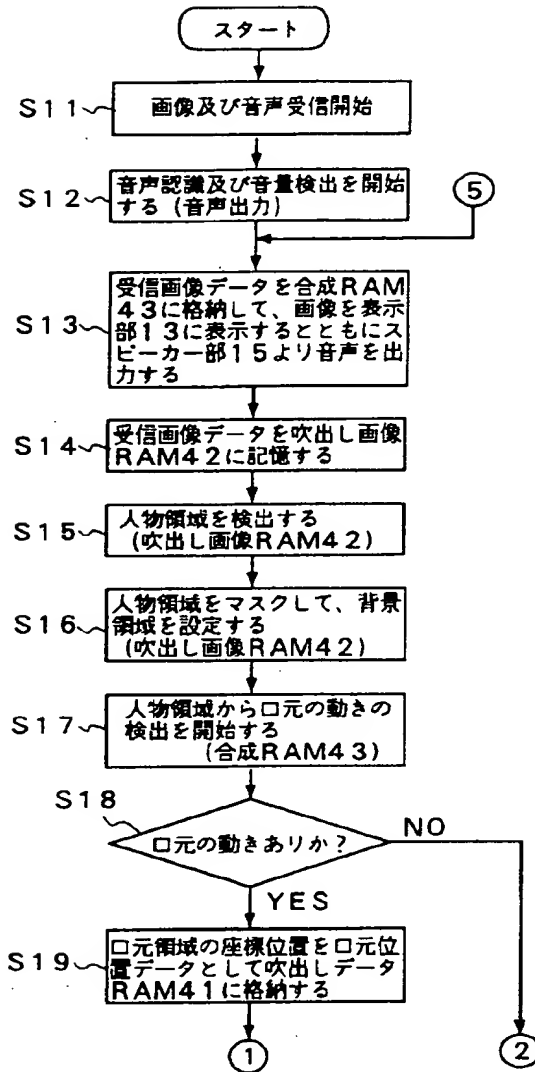
【図11】



【図12】



【図13】



【図14】

